Big Data and Computing Visions

www.bidacv.com

Big. Data. Comp. Vis. Vol. 4, No. 4 (2024) 307–313.

Paper Type: Original Article

An Improvement on the Efficiency Bounds and Efficiency Classifications in DEA with Imprecise Data

Bohlool Ebrahimi[1,*], Duško Tešić[2]

[1] Department of Industrial Engineering, ACECR, Sharif Unit, Tehran, Iran; b.ebrahimi@aut.ac.ir.
[2] Military Academy, University of Defence, Belgrade, Serbia; tesic.dusko@yahoo.com.

Citation:

Received: 03 Januray 2024 *Revised: 06 March 2024* *Accepted: 13 May 2024*	Ebrahimi, B, & Tešić, D. (2024). An improvement on the efficiency bounds and efficiency classifications in DEA with imprecise data. *Big data and computing visions*, 4(4), 307-313.

Abstract

Recently, Park [1] proposed a mathematical Data Envelopment Analysis (DEA) model to estimate the lower bound of efficiency scores in the presence of imprecise data. The current paper shows that its model uses infeasible precise data instead of ordinal data. In addition, in some cases, we may be unable to calculate the relative efficiencies with his model. To overcome the problems, we propose a simple, practical algorithm to estimate the expected value of efficiencies, which is inspired by considering the DEA axioms to the imprecise data.

Keywords: Data envelopment analysis, Efficiency measure, Imprecise data, Ranking.

1 | Introduction

Charnes et al. [2] proposed a Data Envelopment Analysis (DEA) model for the performance evaluation of several similar Decision-Making Units (DMUs) that use multiple inputs to produce multiple outputs. In this model, it is supposed that the values of inputs and outputs are exactly known. However, in real applications, the exact values of inputs and outputs may not be available or cannot be exactly measured. So, these values are sometimes imprecise.

Imprecise data can be of various types, ordinal data (weak and strong), bounded (interval) data, ratio-bound data, multiplied order data, and so on. The mathematical form of this data is presented in [1]. Until now, different approaches have been developed to calculate the relative efficiencies with imprecise data.

For the first time, Cooper et al. [3] used the bounded (interval) and weak ordinal data in DEA and named the new model Imprecise DEA (IDEA), which was a nonlinear and non-convex model. They used the unit-

Corresponding Author: b.ebrahimi@aut.ac.ir
https://doi.org/10.22105/bdcv.2024.486910.1216

invariant property of the DEA model and converted the nonlinear model into an equivalent linear model through scale transformation and variable alterations.

Cooper et al. [4] removed the second problem by introducing some dummy variables. Cooper et al. [5] applied the IDEA approach to evaluate the Korean Mobile Telecommunication Company. Kim et al. [6] used IDEA for performance evaluation in Telephone offices. Lee et al. [7] extended the IDEA concept to the additive DEA model.

Despotis and Smirlis [8] developed two linear programs to estimate the lower and upper-bound efficiencies by considering the pessimistic and optimistic state of each DMU. In fact, the efficiency score for each DMU is an interval in their method. They categorized DMUs into three groups: efficient (the lower bound efficiency score is equal to one), weak efficient (the upper bound efficiency score is equal to one, but the lower bound is less than one) and inefficient (the upper bound efficiency score is less than one).

Zhu [9] showed that the scale transformation (normalization of data) in Cooper et al. [3–5], method is redundant. He converted the bounded and ordinal data into the exact data. He showed that the efficiency score for the exact data is equivalent to the result of solving the nonlinear IDEA model. The proposed method by Zhu [9] reduced the volume of calculations of Cooper et al. [3] method. Zhu [10] used the method for performance evaluation in the Korean Mobile Telecommunication Company. Park [11] reduced the volume of calculations of Cooper et al. [4] method by using a simple variable alteration, which was the same as the variable alteration proposed by Zhu [9].

Wang et al. [12] proposed two linear mathematical programming to obtain the lower bound and upper bound efficiencies by considering a unique (fixed) production frontier for all DMUs.

Kao [13] emphasized that the efficiency scores should be imprecise in the presence of imprecise data. He proposed two mathematical programs to calculate the lower bound and upper bound efficiencies in the presence of weak ordinal and bounded data. Park [1] used the concept of supremum and infimum and proposed a mathematical programming for calculating the lower bound efficiencies.

Park [14] investigated the dual model of IDEA and its relationships with primal problems based on the duality theory in IDEA and developed a computational method for it. Marbini et al. [15] investigated the performance evaluation in the presence of interval data without sign restrictions. Chen et al. [16] presented some models to deal with Likert scale data discrete and bounded data in DEA. They have used the developed DEA models to evaluate the regional energy efficiency in China.

2 | The Problems of Park's Method

Park [1] used the concept of supremum and infimum and proposed the following *Mathematical Programming (1)* for calculating the lower bound efficiencies (DMU_p is under evaluation).

$$
\begin{aligned}
&\max \ \sum_{r=1}^{m} u_r \inf\{y_{rp} \,|\, y_r \in D_r^+\},\\
&\text{s.t.}\\
&\quad \sum_{i=1}^{n} v_i \sup\{x_{ip} \,|\, x_i \in D_i^-\} = 1,\\
&\quad \sum_{r=1}^{m} u_r \inf\{y_{rp} \,|\, y_r \in D_r^+\} - \sum_{i=1}^{n} v_i \sup\{x_{ip} \,|\, x_i \in D_i^-\} \le 0,\\
&\quad \sum_{r=1}^{m} u_r \sup\{y_{rj} \,|\, y_r \in D_r^+\} - \sum_{i=1}^{n} v_i \inf\{x_{ij} \,|\, x_i \in D_i^-\} \le 0,\ j = 1,\dots,k,\ j \ne p,\\
&\quad u_r, v_i \ge \varepsilon,\ \text{for all } r, i.
\end{aligned} \tag{1}
$$

In this model D_i^- and D_r^+ represent the imprecise data. Inf and sup can be replaced with min and max, respectively.

Suppose in *Model (1)*, we have an ordinal input (or output) as follows:

$$x_{i1} \le x_{i2} \le ... \le x_{i,p-1} \le x_{ip} \le x_{i,p+1} \le ... \le x_{in}. \quad \textbf{(2)}$$

At the optimal solution of *Model (1)*, the following relation between the ordinal data should be held.

$$x_{i1}^* \le x_{i2}^* \le ... \le x_{i,p-1}^* \le x_{ip}^* \le x_{i,p+1}^* \le ... \le x_{in}^*. \quad \textbf{(3)}$$

But, in *Model (1)*, the feasibility condition is not considered in the calculation of the Inf and sup by Park [1]. In other words, by using the Park method, at the optimal solution of *Model (1)*, the *Relation (3)* is not established. To clarify the problem, consider the numerical example used in [1]. In this example, there are 8 telephone offices with three inputs and three outputs. The third output is in the ordinal format as follows:

$$D_3^- = \{x_3 \in R^8 \mid x_{34} \ge x_{35} \ge x_{33} \ge x_{37} \ge x_{31} \ge x_{36} \ge x_{32} \ge x_{38}\}. \quad \textbf{(4)}$$

According to the Park method, after normalization, we have

$$D_3^{'-} = \{x_3^{'} \in R^8 \mid 1 \ge x_{34}^{'} \ge x_{35}^{'} \ge x_{33}^{'} \ge x_{37}^{'} \ge x_{31}^{'} \ge x_{36}^{'} \ge x_{32}^{'} \ge x_{38}^{'} \ge 0\}. \quad \textbf{(5)}$$

According to the Park method, when DMU_1 is under evaluation (calculating the lower bound of efficiency), the data for the DMU_1 and other DMUs can be calculated as follows:

$$\begin{aligned} &\sup\{x_{31}^{'} \mid x_3^{'} \in D_3^{'-}\} = \max\{x_{31}^{'} \mid x_3^{'} \in D_3^{'-}\} = 1, \\ &\inf\{x_{3j}^{'} \mid x_3^{'} \in D_3^{'-}\} = \min\{x_{3j}^{'} \mid x_3^{'} \in D_3^{'-}\} = 0; \ \ j = 2,3,...,8. \end{aligned} \quad \textbf{(6)}$$

Therefore, to calculate the lower bound efficiency score of DMU_1, Park [1] has used $x_3^* = (1,0,0,0,0,0,0,0)$, for ordinal data, which is an infeasible solution. As mentioned, the feasibility condition should be considered in the calculation of inf and sup. Applying the feasibility condition implies that he should uses $x_3^* = (1,0,1,1,1,0,1,0)$. Furthermore, Park's method uses only zero and one for all ordinal data. It should be noted that the probability of occurrence of these data is near zero in practice.

The next example shows that in some cases, the Park method is unable to calculate the lower bound efficiencies.

Example 1. Consider three DMUs; each uses one ordinal input to produce one precise output.

Table 1. Data for 3 DMUs.

DMU No.	Input (Ordinal)*	Output (Exact)
1	x_{11}	5
2	x_{12}	3
3	x_{13}	7

* ranking such that $x_{11} \le x_{12} \le x_{13}$.

To calculate the lower bound of the efficiency score of these three DMUs, the Park method uses only zero and one for the ordinal data, as presented in *Table 2*.

Table 2. The values of ordinal data in the calculation of lower bound efficiency with Park's method.

The DMU under Evaluation / The Values of Variables	Without Considering Feasibility			Considering Feasibility		
	DMU_1	DMU_2	DMU_3	DMU_1	DMU_2	DMU_3
x'^*_{11}	1	0	0	1	0	0
x'^*_{12}	0	1	0	1	1	0
x'^*_{13}	0	0	1	1	1	1
Lower bound efficiencies	Cannot be calculated	Cannot be calculated	Cannot be calculated	0.17	Cannot be calculated	Cannot be calculated

As can be seen from *Table 2*, some DMUs produce output without consuming any input. With this data, the efficiencies cannot be calculated, and so the ranking of DMUs is not possible.

3 | The Proposed Algorithm

In the following, a simple algorithm is presented to estimate the relative efficiencies. The algorithm is inspired by considering the DEA axioms in the presence of imprecise data.

It should be emphasized that the CCR model and Production Possibility Set (PPS) are based on some axioms. Suppose we have k DMUs, (DMU_j, $j=1,2,3,...,k$) with n inputs ($x_{ij} \geq 0$, $i=1,2,3,...,n$; $j=1,2,3,...,k$) and m outputs (y_{rj}, $r=1,2,3,...,m$; $j=1,2,3,...,k$), such that at least one input and one output is nonzero for each DMU. The axioms are as follows:

I. Inclusion of observation: each observed DMU_j belongs to T, (j=1,2,…,k). T is the PPS.

II. Free disposability of inputs and outputs: if $(x,y) \in T, y' \leq y,$ then $(x,y') \in T$ and if $(x,y) \in T, x' \geq x,$ then $(x',y) \in T$.

III. Convexity: if $(x,y) \& (x',y') \in T \Rightarrow \lambda(x,y)+(1-\lambda)(x',y') \in T,$ for all $0 \leq \lambda \leq 1$.

IV. Constant returns to scale: if $(x,y) \in T \Rightarrow (\lambda x, \lambda y) \in T,$ for all $0 \geq \lambda$.

V. Minimum extrapolation: T is the intersection of all sets satisfying 1-4.

When the data are imprecise, in fact, there is not any observed data, and so the first axiom (Inclusion of observation) is not established. In this situation, we cannot build the PPS correctly. To clarify the topic, consider the data presented in *Table 3*.

Table 3. Data for 2 DMUs.

DMU no.	Input 1	Output 1*
1	1	y_1
2	2	y_2

*which $y_1 \geq y_2$.

For this data, the PPS and the production frontier with constant returns to scale technology have been shown in *Fig. 1*. Indeed, the production frontier for this PPS is the line between $(0,0)$ and $(1,+\infty)$, that is unacceptable. This fault shows that observing the DMU data (first axiom) is necessary, and the PPS and production frontier could not be built correctly without considering the axiom.

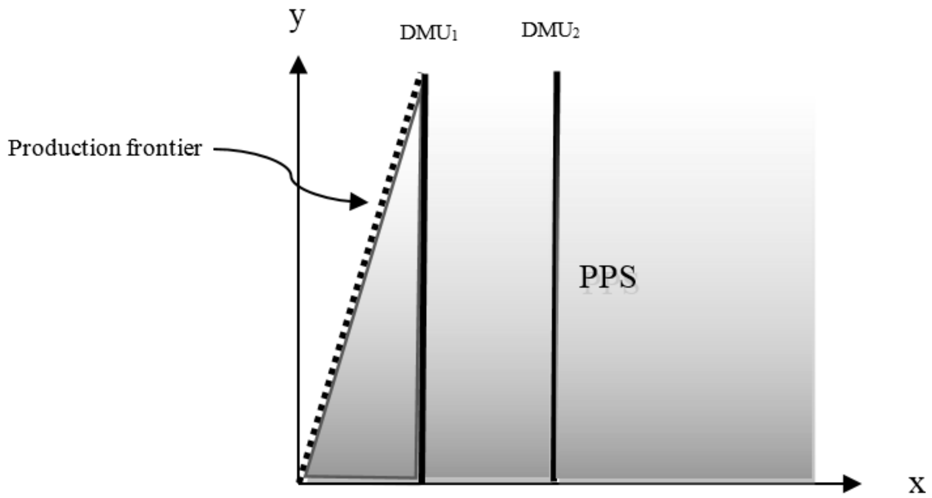

Fig. 1. PPS for the data presented in table 3.

For considering the DEA axiom, the imprecise data can be regarded as stochastic data. Because the probability distribution of this data is unknown, it can be supposed to be uniform. In this case, the steps of the proposed algorithm are as follows:

I. Generating data with the uniform distribution (by considering the relations among ordinal data) with *N* iterations.

II. Calculating efficiencies for all DMUs with the data obtained from Stage 1. At this stage, *N* efficiency scores for each DMU will be obtained.

III. Calculating the average efficiency score and standard deviation of efficiencies. Indeed, the average efficiency is an estimation of the expected value of efficiencies.

IV. Ranking DMUs based on average efficiency scores and standard deviation of efficiencies as follows DMU_{j1} dominates DMU_{j2} if and only if the average efficiency score of DMU_{j1} is greater than DMU_{j2}. If two DMUs have the same average efficiency score, the DMU with less standard deviation has a better rank.

Obviously, increasing N leads to obtaining a more exact average efficiency. If we were able to extract the efficiency distribution, we could calculate the expected value of relative efficiency for each DMU. Indeed, the average efficiency is an estimation of expected relative efficiency, so these values will be approximately equal by increasing the value of N.

4 | Numerical Example

In this section, we apply the proposed approach in this paper to the efficiency measure of five DMUs studied by Cooper et al. [3]. The DMUs have two inputs (one exact and one interval) and two outputs (one exact and one ordinal), as presented in *Table 4*.

Table 4. The values of inputs and outputs for 5 DMUs.

DMUs No.	Inputs		Outputs		Park (2007)	Ranking
	x_{1j} (Exact)	x_{2j} (Interval)	y_{1j} (Exact)	y_{2j} (Ordinal)*		
1	100	[0.6, 0.7]	2000	4	[1, 1]	1
2	150	[0.8, 0.9]	1000	2	[0.33333, 0.87499]	3
3	150	1	1200	5	[0.4, 1]	2
4	200	[0.7, 0.8]	900	1	[0.33748, 0.99999]	3
5	200	1	600	3	[0.17999, 0.69999]	3

*ranking such that: $y_{23} \geq y_{21} \geq y_{25} \geq y_{22} \geq y_{24}$.

The results of Park's method are summarized in the last two columns of *Table 4* with $\varepsilon = 10^{-6}$. Now, we apply the proposed method in this paper to rank these five DMUs. We generated $N = 1, 10, 100, 1000, 5000$ and

10000 random data for ordinal and interval data and calculated the average efficiency score and standard deviation for these DMUs. The results are summarized in *Table 5*. As can be seen, by increasing the value of N, the variations of average efficiencies and standard deviations decrease. The variations are less than 0.01. The results show that generating $N \geq 1000$ random data can lead to a reliable ranking. The final rank based on 10000 data generations is given in the last row of *Table 5*. As can be seen, the ranking is as follows: DMU_1> DMU_3> DMU_2> DMU_4> DMU_5.

Table 5. The results of the proposed algorithm for data presented in table 2.

N	Average Efficiency Scores and Standard Deviation									
	DMU_1		DMU_2		DMU_3		DMU_4		DMU_5	
	A.E.S*	S.D*	A.E.S	S.D	A.E.S	S.D	A.E.S	S.D	A.E.S	S.D
1	1	0	0.4179	0	1	0	0.3635	0	0.4519	0
10	1	0	0.4014	0.0787	0.8922	0.1409	0.3837	0.0245	0.3478	0.1325
100	1	0	0.3918	0.0338	0.9351	0.1023	0.3939	0.0357	0.2846	0.1234
1000	1	0	0.3952	0.0478	0.9414	0.1014	0.3939	0.0305	0.2947	0.1253
5000	1	0	0.3944	0.0448	0.9403	0.1020	0.3937	0.0328	0.2945	0.1253
10000	1	0	0.3950	0.0495	0.9401	0.1034	0.3933	0.0417	0.2941	0.1236
Rank	1		3		2		4		5	

* A.E.S: average efficiency score; S.D: standard deviation.

5 | Conclusion

The paper shows that the proposed model by Park [1] has two problems, infeasibility problem and being unable to calculate the relative efficiencies. It was emphasized that the DEA model and PPS are based on some axioms, especially the inclusion of the observation axiom. It is shown that with imprecise data, we may be unable to determine the PPS and the production frontier correctly.

To overcome the drawbacks and also to consider the DEA axioms, we proposed a new method based on data generation to estimate the efficiency score in the presence of imprecise data. We considered the average efficiency and standard deviation for ranking DMUs. It was shown that this method generates more realistic and reliable results.

Author Contributions

Bohlool Ebrahimi and Duško Tešić equally contributed to the conceptualization and methodology of this study. Bohlool Ebrahimi identified the limitations in Park's DEA model and proposed the algorithm to estimate the expected value of efficiencies, while Duško Tešić conducted the theoretical validation and practical implementation of the algorithm. Both authors participated in the writing, editing, and review of the manuscript.

Funding

No specific funding was received for this research.

Data Availability

This study is based on theoretical models and simulations. All necessary data and computational steps are provided within the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding this research

References

[1] Park, K. S. (2007). Efficiency bounds and efficiency classifications in DEA with imprecise data. *Journal of the operational research society*, *58*(4), 533–540. https://www.tandfonline.com/doi/abs/10.1057/palgrave.jors.2602178

[2] Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, *2*(6), 429–444. https://www.sciencedirect.com/science/article/pii/0377221778901388

[3] Cooper, W. W., Park, K. S., & Yu, G. (1999). IDEA and AR-IDEA: models for dealing with imprecise data in DEA. *Management science*, *45*(4), 597–607. https://pubsonline.informs.org/doi/abs/10.1287/mnsc.45.4.597

[4] Cooper, W. W., Park, K. S., & Yu, G. (2001). IDEA (imprecise data envelopment analysis) with CMDs (column maximum decision making units). *Journal of the operational research society*, *52*(2), 176–181. DOI:10.1057/palgrave.jors.2601070

[5] Cooper, W. W., Park, K. S., & Yu, G. (2001). An illustrative application of IDEA (imprecise data envelopment analysis) to a Korean mobile telecommunication company. *Operations research*, *49*(6), 807–820.

[6] Kim, S. H., Park, C. G., & Park, K. S. (1999). An application of data envelopment analysis in telephone officesevaluation with partial data. *Computers & operations research*, *26*(1), 59–72. DOI:https://doi.org/10.1016/S0305-0548(98)00041-0

[7] Lee, Y. K., Sam Park, K., & Kim, S. H. (2002). Identification of inefficiencies in an additive model based IDEA (imprecise data envelopment analysis). *Computers & operations research*, *29*(12), 1661–1676. https://www.sciencedirect.com/science/article/pii/S0305054801000491

[8] Despotis, D. K., & Smirlis, Y. G. (2002). Data envelopment analysis with imprecise data. *European journal of operational research*, *140*(1), 24–36. https://www.sciencedirect.com/science/article/pii/S0377221701002004

[9] Zhu, J. (2003). Imprecise data envelopment analysis (IDEA): a review and improvement with an application. *European journal of operational research*, *144*(3), 513–529. https://www.sciencedirect.com/science/article/pii/S0377221701003927

[10] Zhu, J. (2004). Imprecise DEA via standard linear DEA models with a revisit to a Korean mobile telecommunication company. *Operations research*, *52*(2), 323–329. https://pubsonline.informs.org/doi/abs/10.1287/opre.1030.0072

[11] Park, K. S. (2004). Simplification of the transformations and redundancy of assurance regions in IDEA (imprecise DEA). *Journal of the operational research society*, *55*(12), 1363–1366.

[12] Wang, Y. M., Greatbanks, R., & Yang, J. B. (2005). Interval efficiency assessment using data envelopment analysis. *Fuzzy sets and systems*, *153*(3), 347–370. https://www.sciencedirect.com/science/article/pii/S0165011404005767

[13] Kao, C. (2006). Interval efficiency measures in data envelopment analysis with imprecise data. *European journal of operational research*, *174*(2), 1087–1099. https://www.sciencedirect.com/science/article/pii/S037722170500319X

[14] Park, K. S. (2010). Duality, efficiency computations and interpretations in imprecise DEA. *European journal of operational research*, *200*(1), 289–296. https://www.sciencedirect.com/science/article/pii/S0377221708010308

[15] Hatami-Marbini, A., Emrouznejad, A., & Agrell, P. J. (2014). Interval data without sign restrictions in DEA. *Applied mathematical modelling*, *38*(7), 2028–2036. https://www.sciencedirect.com/science/article/pii/S0307904X13006392

[16] Chen, Y., Cook, W. D., Du, J., Hu, H., & Zhu, J. (2017). Bounded and discrete data and Likert scales in data envelopment analysis: application to regional energy efficiency in China. *Annals of operations research*, *255*, 347–366. https://link.springer.com/article/10.1007/s10479-015-1827-3